A Distance Weight Matrix Determination for Superficial Spatial Data

Determinazione di una matrice di pesi per dati spaziali

Letizia La Tona
Dip. di Statistica, University of Messina, latona@unime.it

Angelo Mazza
Dip. di Economia e Metodi Quantitativi, University of Catania, a.mazza@unict.it

Massimo Mucciardi
Dip.di Ec. Stat. e Analisi del Territorio, University of Messina, mucciard@unime.it

Riassunto: In questo lavoro presentiamo un criterio per la determinazione di una matrice di pesi per dati spaziali di tipo areale che fornisce una soluzione al problema dell'invarianza topologica. Il criterio proposto rende possibile, sulla base della distanza tra centroidi delle configurazioni spaziali osservate, determinare una matrice dei pesi sensibile ai fattori di distanza e di forma delle unità. La elaborazione di un programma in Visual Basic rende la procedura applicabile a situazioni reali.

Keywords: Weight Matrix, Superficial Data, Spatial Autocorrelation

1. Introduction

The majority of "traditional" statistical tools are based on the assumption of independent observations; they are therefore inappropriate for processing data where this assumption is in practice systematically violated. In space, it is unusual to find independence between the various observations; on the contrary, phenomena of an economic, social or environmental nature are characterised precisely by spatial dependence. Moreover, the variables observed for the study of territorial units (La Tona, 2003), or of space in general, are interdependent in all directions and refer to limits established with an often arbitrary partition of the territory, so the lack of a definite determination leads to various possible aggregations. Over the years, various statistics have been created to measure the dependence of spatial data: simple indices such as Moran's "I" statistic or Geary's "C", derived indices such as Ripley's K statistic, multivariate indices in regression models, and so on. The majority of them are based on the so-called system of interconnection, that is on a mechanism able to organise spatial units territorially (a contiguity matrix). The main problem of traditional interconnection systems is that they are invariant with regard to topological transformations of the territory (Dacey, 1968). In the past, various scholars have tried to solve this problem by introducing distance functions within the contiguity matrix, in order to take into consideration certain physical elements which play a more or less important role in spatial organisation. Cases in point are Dacey (1968), who considers the form of the units, and Cliff and Ord (1981), who introduce measurements of distance and the length of the borders between spatial units. More recently, the criterion of a “Double State Maximin Algorithm” (DSMA, Mucciardi, 1998) has been proposed, which takes into consideration the effective distance between spatial units represented as points, and returns final weights proportional to the intensity of the links between interconnected units. This paper represents an extension of the DSMA procedure capable of: 1) determining beforehand the territorial partitions between the units on the basis of a mechanism based on distances between territorial barycentres or centroids; 2) establishing, within the generic territorial partition, a function that takes into consideration the distance and the physical characteristics of the superficial units, with the aim of obtaining a spatial organisation in relation to the characteristics mentioned; 3) obtaining, by means of the implemented software, output compatible with the measurement of spatial autocorrelation. The procedure proposed has the important property of being “sensitive” to topological transformations of the territory. In the following paragraphs we will expose the procedure and give for a simulated example the numerical results both of the proposed method DSMA and of the classical method.

2. The “Double State Maximin Algorithm” criterion and its extension

The DSMA procedure automatically determines weight matrices whose ε_{ij}^{k} values represent the interconnection, in the generic partition k, between spatial units represented as points. The weights generated in the “interconnection ray” h^k are sensitive to the effective distance of each unit. In other words, let us suppose k as the generic partition and $\mathbf{T}^k$ as a matrix with values:

$$t_{ij}^{k} = \left(\frac{1}{\lambda^k} - \frac{Max\left(\mathbf{G^k}\right) - g_{ij}^{k}}{Max\left(\mathbf{G^k}\right)} \right) \text{ if } d^{k}{}_{ij} \leq h^k \text{ ; } t_{ij}^{k} = 0 \text{ if } d^{k}{}_{ij} > h^k \qquad (1)$$

where $g^{k}{}_{ij} = h^k - d^{k}{}_{ij}$ represents the difference between the interconnection ray h^k and the distance $d^{k}{}_{ij}$, while λ^k is a classification index (Mucciardi, 1997-1998) which provides information on the usefulness of the partitions generated. In particular, λ^k is equal to zero only in theoretical spatial arrangements, such as equidistant units where $h^k = d^{k}{}_{ij}$. To ensure the compatibility of the elements of the matrix $\mathbf{T}^k$ with the measurements of spatial autocorrelation (interconnections of partition k), we introduce a reweighting of the matrix $\mathbf{T}^k$ such that:

$$\varepsilon_{ij}^{k} = \alpha^k \cdot t_{ij}^{k} \quad \text{with}\, \alpha^k = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^{k}}{\sum_{i=1}^{n}\sum_{j=1}^{n} t_{ij}^{k}} \qquad (2)$$

where ε_{ij}^{k} is the generic element of the reweighted matrix $\mathrm{E}^{\mathbf{k}}$. (The reweighting ensures that $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^{k} = \sum_{i=1}^{n}\sum_{j=1}^{n} \varepsilon_{ij}^{k}$). Let us suppose $\mathbf{r_i}$ with $i = 1....n$, a vector which embodies a certain physical feature of the n areal units (e.g. measurement of surface). From the vector $\mathbf{r_i}$ we generate a matrix $\mathbf{Q}^k$ with a generic element:

$$q_{ij}^{k} = \left(\frac{M}{\sigma} - \frac{f_{ij}^{k} - Max\left(f_{ij}^{k}\right)}{Max\left(f_{ij}^{k}\right)} \right) \text{ if } d^{k}{}_{ij} \le h^{k}\text{; } q_{ij}^{k} = 0 \text{ if } d^{k}{}_{ij} > h^{k} \tag{3}$$

where $f_{ij}^{k} = \left| r_i - r_j \right|$ is a measurement of the inequality between the units i and j in the partition k, while M and σ represent respectively the mean and the average quadratic loss of the vector $\mathbf{r_i}$. It is easy to verify that the quantities q_{ij}^{k} are included between:

$$\frac{M}{\sigma} \le q_{ij}^{k} \le \frac{M}{\sigma} + 1 \tag{4}$$

When, respectively, $f_{ij}^{k} = Max(f_{ij}^{k})$ and $f_{ij}^{k} = 0$. Also in this case, to ensure the compatibility of the elements of the matrix $\mathbf{Q}^k$ with the measurements of spatial autocorrelation (interconnections of partition k), we reweight matrix $\mathbf{Q}^k$ such that:

$$\gamma_{ij}^{k} = \beta^{k} \cdot q_{ij}^{k} \qquad \text{with } \beta^{k} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^{k}}{\sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij}^{k}} \tag{5}$$

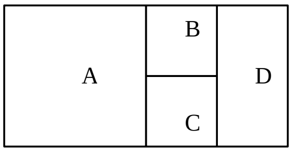
with γ_{ij}^{k} as the generic element of the reweighted matrix $\mathbf{B^k}$ (The reweighting ensures that $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^{k} = \sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij}^{k}$). In the case of areal units which are all perfectly equal (regular grid), we will have: $\sigma \to 0$, $q_{ij}^{k} \to \infty$, $\gamma_{ij}^{k} \to 1$, while, for unequal areal units (irregular grid), we will have: $\sigma > 0$, $\gamma_{ij}^{k} > 1$ if $q_{ij}^{k} > \frac{1}{\beta^{k}}$ $\sigma > 0$, $\gamma_{ij}^{k} < 1$ if $q_{ij}^{k} < \frac{1}{\beta^{k}}$.
Thus, if two territorial units represented as areas present a similar feature (for example, surface), the procedure provides an output greater than one subordinated to the ratio $\frac{M}{\sigma}$. Vice versa, if they present different features, the procedure provides an output lower than one subordinated to the ratio $\frac{M}{\sigma}$.
By introducing a suitable function on the coefficients obtained in the relations (2) and (5), it is possible to obtain a matrix $\mathbf{\Omega^k}$ with a generic element equal to:

$$\omega_{ij}^{k} = f\left(\varepsilon_{ij}^{k}, \gamma_{ij}^{k}\right) \text{ if } d^{k}{}_{ij} \le h^{k} \text{ with } (k = 1 \ldots t) \;\; \omega_{ij}^{k} = 0 \text{ if } d^{k}{}_{ij} > h^{k} \qquad (6)$$

3. Software implemented and obtained results

The procedure may be easily applied through a software developed in Visual Basic to a significantly high number of units, but for lack of space, by way of example, we show a simulated territorial configurations. The software allows us to use any function of the coefficients ε_{ij}^{k} and γ_{ij}^{k} with constraint: $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^{k} = \sum_{i=1}^{n}\sum_{j=1}^{n} \omega_{ij}^{k}$. The output generate final coefficients (comparable to the weights of the contiguity matrices) able to evaluate simultaneously or singularly, in each territorial partition, "distance" and "surface" factors in areal units. We use two simulated territorial configurations, and suppose: 1) territorial barycentre as a matrix of distance; 2) surface as a physical characteristic; 3) the mean of the two coefficients as $f\left(\varepsilon_{ij}^{k}, \gamma_{ij}^{k}\right)$. The application of traditional procedure to the hypothesised situation, as it is known, would produce two coincident binary matrices; vice versa, the D.S.M.A procedure generates two different matrices whose weights are sensitive to the form and the distance characteristics of the units.

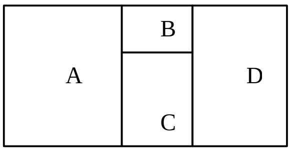
Figure 1: *Weight matrix (first spatial lag) for two simulated territorial configurations: the binary contiguity matrix, and proposed procedure D.S.M.A.*



Area	A	B	C	D
A	0	1	1	0
B	1	0	1	1
C	1	1	0	1
D	0	1	1	0

Area	A	B	C	D
A	0,00	0,82	0,82	0,00
B	0,82	0,00	1,23	1,06
C	0,82	1,23	0,00	1,06
D	0,00	1,06	1,06	0,00

λ
0,27
ΣΣω
10



Area	A	B	C	D
A	0	1	1	0
B	1	0	1	1
C	1	1	0	1
D	0	1	1	0

Area	A	B	C	D
A	0,00	0,89	1,02	0,00
B	0,89	0,00	1,13	0,93
C	1,02	1,13	0,00	1,02
D	0,00	0,93	1,02	0,00

λ
0,17
ΣΣω
10

Traditional procedure   DSMA procedure

References

CLIFF J., ORD K. (1981), *Spatial Process*, Pion Limited, London.

DACEY M. (1968), A review of measures of contiguity for two and k-color maps, in: *Spatial analysis: a reader in spatial geography,* Berry-Marble (Eds.), Prentics-Hall, New York, 479-495.

LA TONA L. (2003), I modelli multilevel per lo studio di dati territoriali, *SIEDS 2003*, Campobasso.

MUCCIARDI M. (1998), La procedura D.S.M.A. per la misura dell'intensità dei legami tra unità spaziali di tipo puntuale, *SIS XXXIX Riunione Scientifica*, Sorrento, 773-779.